Visualizing which parts of IIIF images are looked by users

AUTHOR(S):

Nishioka, Chifumi; Nagasaki, Kiyonori

Visualizing which Parts of IIIF Images are Looked by Users

Chifumi Nishioka, Kyoto University Library

Kiyonori Nagasaki, The University of Tokyo

2018 IIIF Conference, Washington DC, US

京都大学 KYOTO UNIVERSITY

東京大学 THE UNIVERSITY OF TOKYO

Evaluating the Usage of DAs

Evaluating the usage of the digital archives is important

- Evaluation measures for the usage of digital archives
  - Number of hits, pageviews, and visitors
  - Number of accesses to each bibliography
  - Number of accesses to each image

In IIIF, an image is called via IIIF Image API **with specifying a region of an image**

IIIF Image API: {scheme}://{server}{/prefix}/{identifier}/**{region}**/{size}/{rotation}/{quality}.{format}

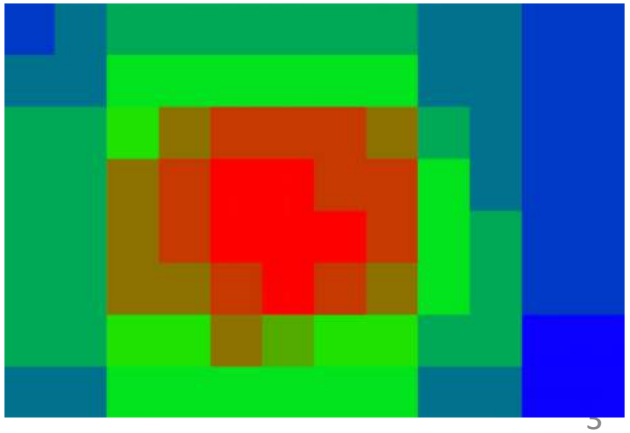
IIIF enables more fine-grained analysis of usage of images

Generating Heatmaps

Analyzing IIIF Image API logs, we generate heatmaps that visualize which parts of IIIF images are looked by users

Python Script

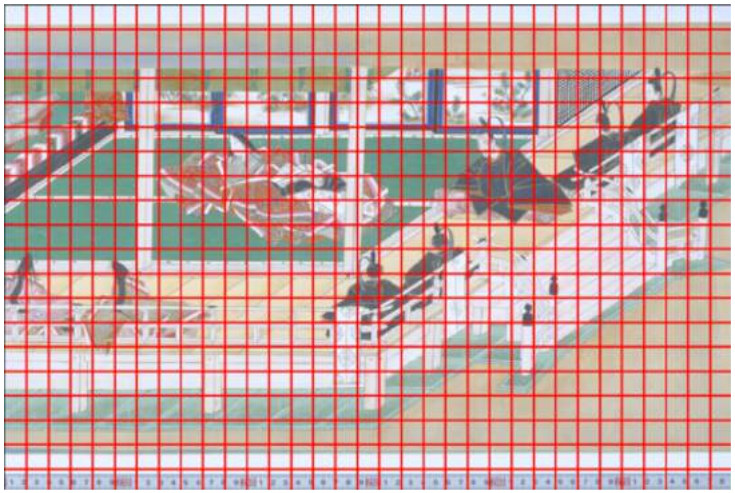
- Prepare H × W matrices for each image
  - H: height of an image, W: width of an image
  - Each element in matrices corresponds to each pixel
  - The size of images is retrieved from info.json
- Count the number of accesses to each pixel and record it to H * W matrix
- Generate heatmaps
  - Calculate RGB values for values in matrices
  - Output matrices as images



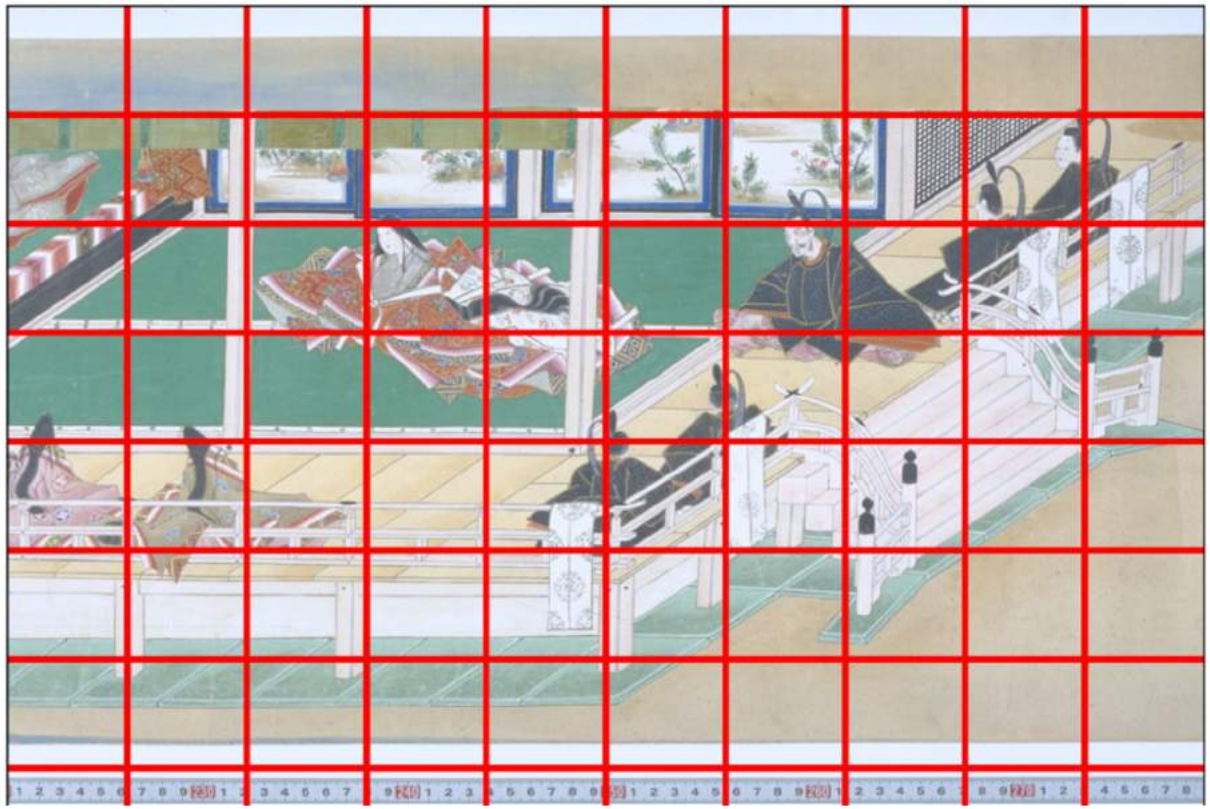
Photograph courtesy of the Main Library, Kyoto University - Tamamo-no-mae adapted

Speed Up

Count the number of accesses to each pixel



Count the number of accesses in N-pixel unit



Computation time for counting # of accesses (100k access logs, in which 27,736 logs are calls of IIIF image API)

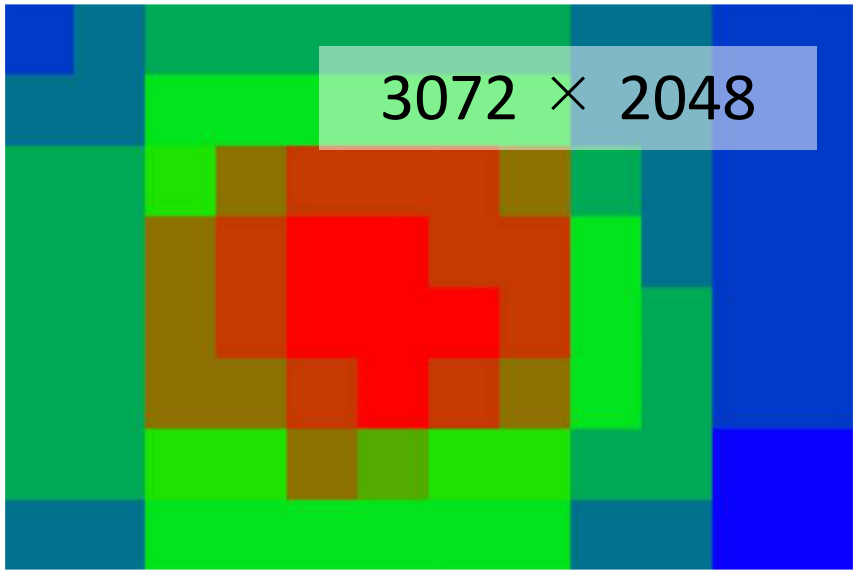
10-pixel unit: 84.23 (s)

100 pixel unit: 1.09 (s)

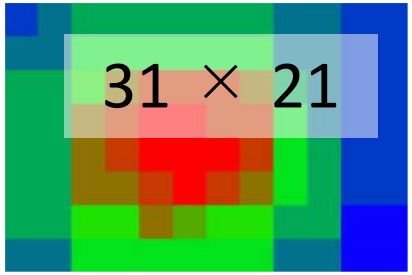
Computer used for the experiment: iMac (macOS High Sierra version 10.13.4), Processor 4GHz Intel Core i7, Memory 16GB, 1867 MHz DDR3

Speed Up Further

Output a heatmap with the size of an original IIIF image



Output in a small size



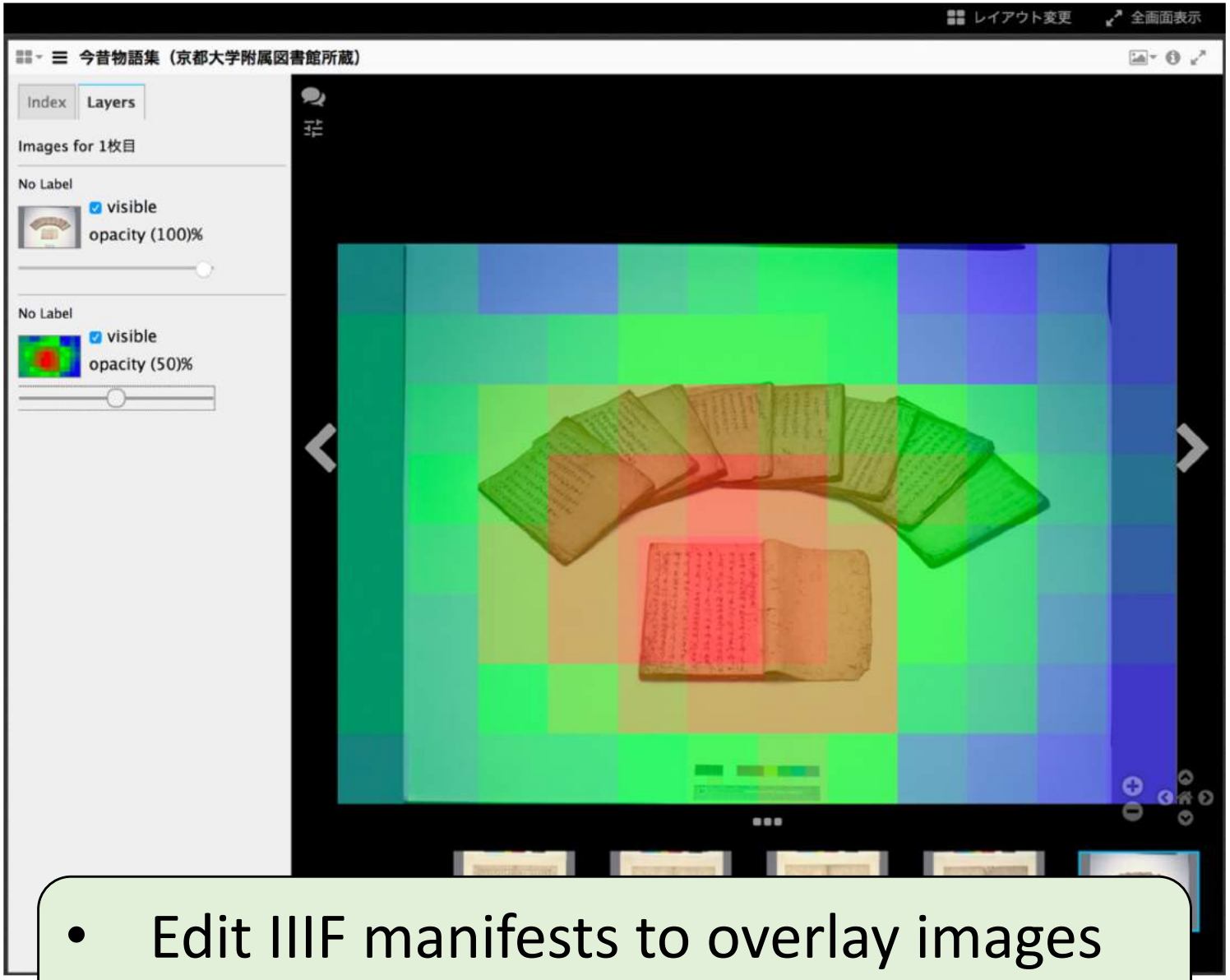
Average computation time for generating one heatmaps (calculating RGB value for each pixel and output as an image)

10-pixel unit: 2.05 (s) (SD: 7.01)

100 pixel unit: 0.02 (s) (SD: 0.07)

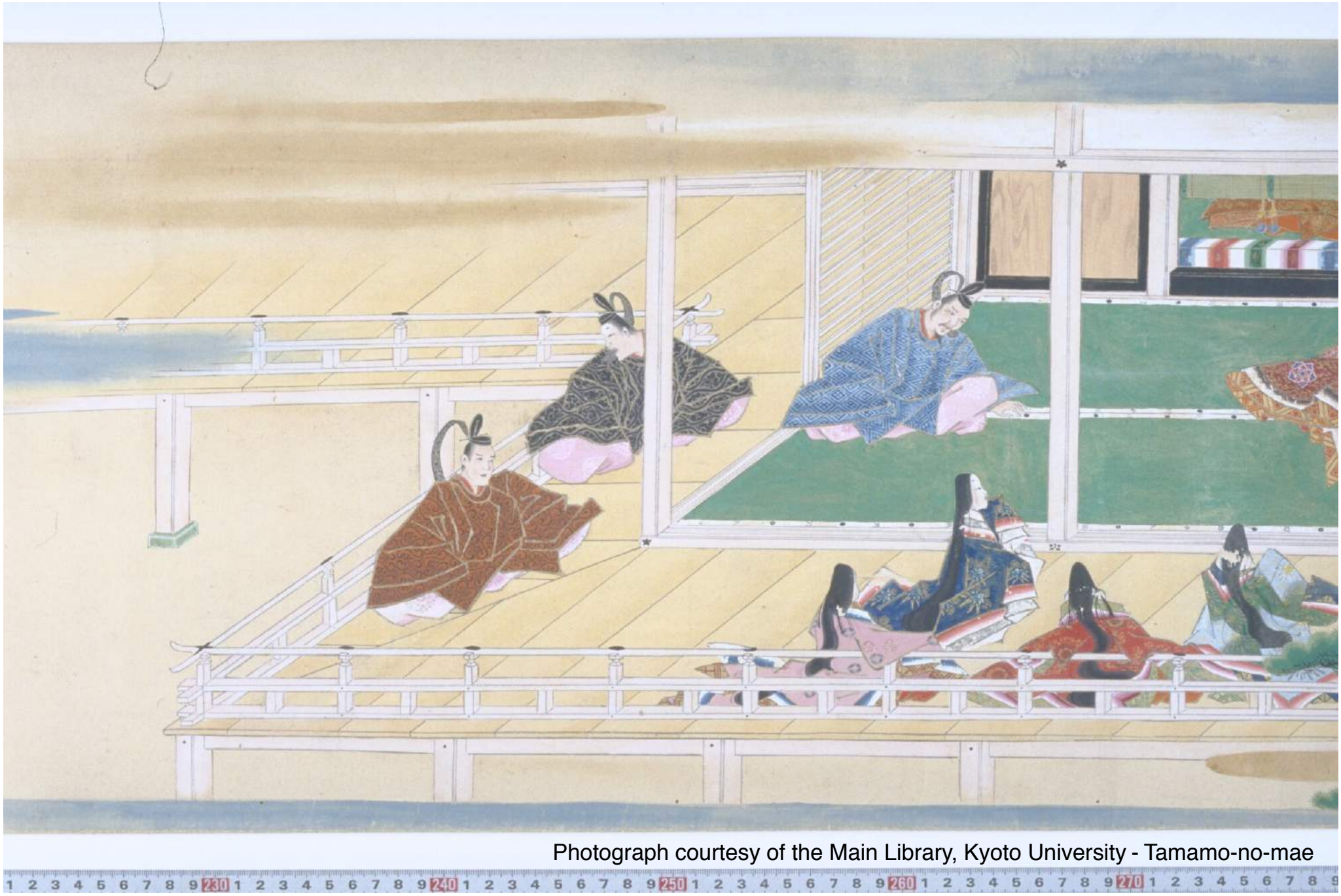
Computer used for the experiment: iMac (macOS High Sierra version 10.13.4), Processor 4GHz Intel Core i7, Memory 16GB, 1867 MHz DDR3

Displaying Heatmaps over Images



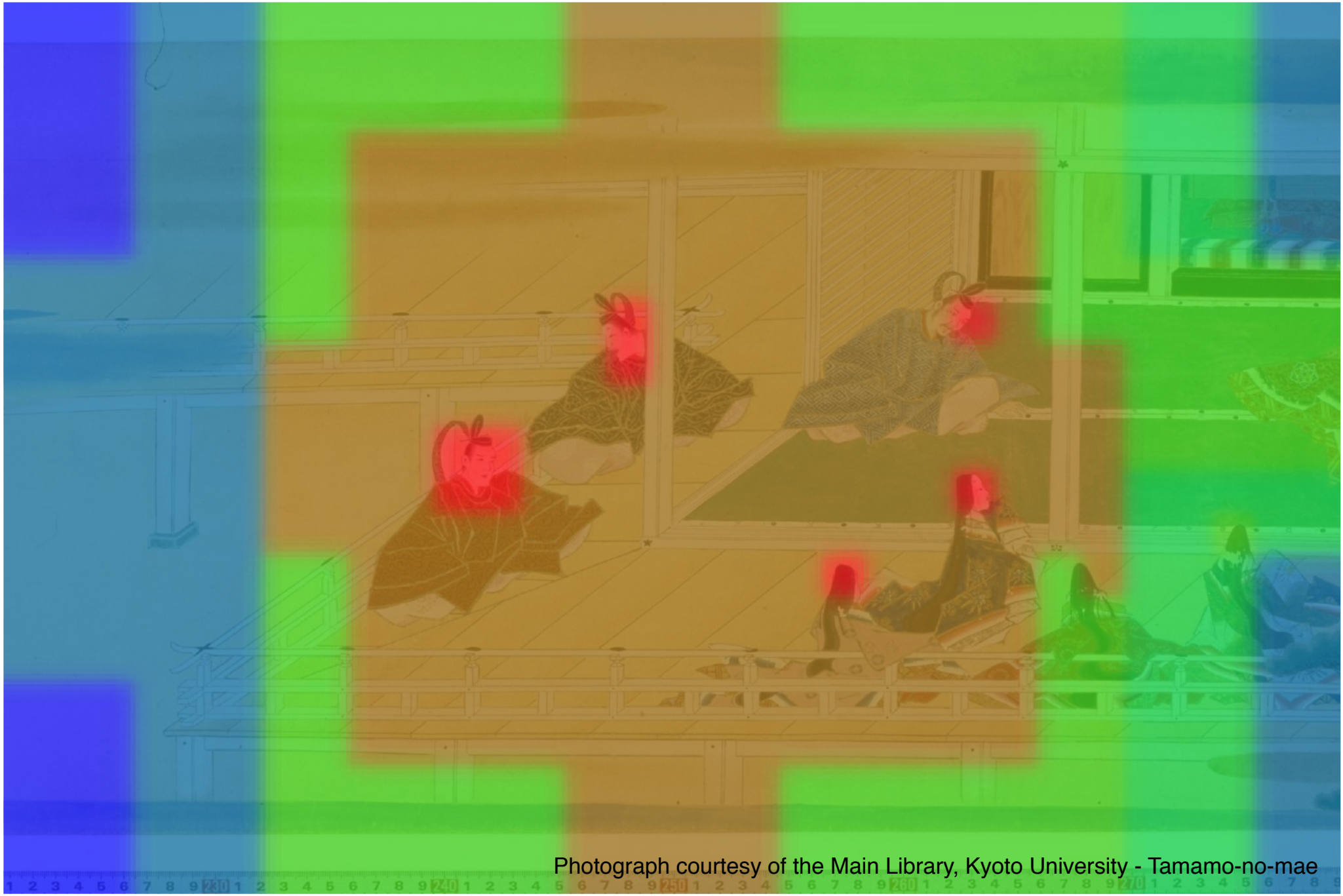
- Edit IIIF manifests to overlay images
- Use Mirador’s layer function

Example



Photograph courtesy of the Main Library, Kyoto University - Tamamo-no-mae

Example



Photograph courtesy of the Main Library, Kyoto University - Tamamo-no-mae

Possible Applications

Thumbnails

- Most-viewed regions of images are used as thumbnails.

Research Collaborations

- Collaborators can see which parts of images have been already investigated.
- A tool to stimulate motivation for crowd-sourcing

Understanding research process

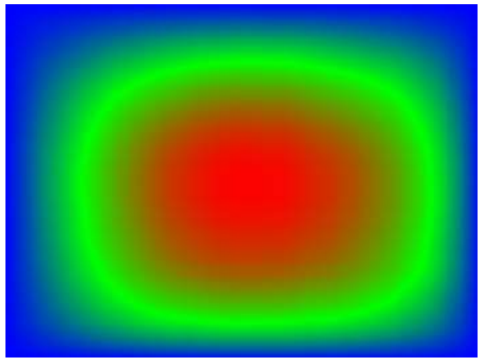
- Researchers can reflect their own research process.
- Young researchers can learn research methodology by looking how experienced researchers do their research.

Risks and Concerns

- Visualization of access logs is not a problem, if anonymization is conducted appropriately
- However, anonymization can be invalidated for IIIF images with few accesses
  - In the field where a small number of researchers work, peers can easily guess who accessed and investigated images
- In addition, a series of activities on IIIF images might reveal his/her viewpoint that would be a key issue of his/her academic outcome
  - Key issues can be revealed even before publication of outcome
  - Priority rights of research can be spoiled
- Therefore, we need a careful management of access logs to make services for researchers trustworthy

Future Works

- Investigate risks and concerns carefully
  - How many accesses do we need to ensure that anonymization cannot be invalidated?
  - Formulate a guideline of management and usage of access logs
- Real-time processing (i.e., stream processing)
  - How to update heatmaps as they get new accesses
- Take probabilities of being accessed of different regions into consideration
  - Regions close to the center of images have higher probability to be accessed
  - Should we reduce counts of regions close to the center when generating heatmaps?
  - It might reveal interesting insights…



Thank you!